Arthritis Research & Therapy

RESEARCH ARTICLE Open Access

High-throughput sequencing of CD4$^+$ T cell repertoire reveals disease-specific signatures in IgG4-related disease

Liwen Wang[1,2†], Panpan Zhang[1†], Jieqiong Li[1], Hui Lu[1], Linyi Peng[1], Jing Ling[3], Xuan Zhang[1], Xiaofeng Zeng[1], Yan Zhao[1*] and Wen Zhang[1*]

Abstract

Background: CD4$^+$ T cells play critical roles in the pathogenesis of IgG4-related disease (IgG4-RD). The aim of this study was to investigate the TCR repertoire of peripheral blood CD4$^+$ T cells in IgG4-RD.

Methods: The peripheral blood was collected from six healthy controls and eight IgG4-RD patients. TCR β-chain libraries of CD4$^+$ T cells were constructed by 5′-rapid amplification of cDNA ends (5′-RACE) and sequenced by Illumina Miseq platform. The relative similarity of TCR repertoires between samples was evaluated according to the total frequencies of shared clonotypes (metric F), correlation of frequencies of shared clonotypes (metric R), and total number of shared clonotypes (metric D).

Results: The clonal expansion and diversity of CD4$^+$ T cell repertoire were comparable between healthy controls and IgG4-RD patients, while the proportion of expanded and coding degenerated clones, as an indicator of antigen-driven clonal expansion, was significantly higher in IgG4-RD patients. There was no significant difference in TRBV and TRBJ gene usage between healthy controls and IgG4-RD patients. The complementarity determining region 3 (CDR3) length distribution was skewed towards longer fragments in IgG4-RD. Visualization of relative similarity of TCR repertoires by multi-dimensional scaling analysis showed that TCR repertoires of IgG4-RD patients were separated from that of healthy controls in F and D metrics. We identified 11 IgG4-RD-specific CDR3 amino acid sequences that were expanded in at least 2 IgG4-RD patients, while not detected in healthy controls. According to TCR clonotype networks constructed by connecting all the CDR3 sequences with a Levenshtein distance of 1, 3 IgG4-RD-specific clusters were identified. We annotated the TCR sequences with known antigen specificity according to McPAS-TCR database and found that the frequencies of TCR sequences associated with each disease or immune function were comparable between healthy controls and IgG4-RD patients.

Conclusion: According to our study of CD4$^+$ T cells from eight IgG4-RD patients, TCR repertoires of IgG4-RD patients were different from that of healthy controls in the proportion of expanded and coding degenerated clones and CDR3 length distribution. In addition, IgG4-RD-specific TCR sequences and clusters were identified in our study.

Keywords: IgG4-related disease, CD4$^+$ T cells, TCR repertoire, Complementarity determining region 3, Antigen

* Correspondence: zhaoyan_pumch2002@aliyun.com; zhangwen91@sina.com
[†]Liwen Wang and Panpan Zhang contributed equally to this work.
[1]Department of Rheumatology, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Key Laboratory of Rheumatology and Clinical Immunology, Ministry of Education, No.41 Da Mu Cang, Western District, Beijing 100032, People's Republic of China
Full list of author information is available at the end of the article

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

IgG4-related disease (IgG4-RD) is a newly recognized clinical entity mainly affecting middle-aged to elderly males, characterized by immune-mediated fibro-inflammatory process. Pathologic features of IgG4-RD include dense lymphoplasmacytic infiltration enriched in IgG4-positive plasma cells, storiform fibrosis, and obliterative phlebitis [1, 2]. The pathogenesis of IgG4-RD remains unclear. It has been proposed that chronic antigen stimulation induces activation, clonal expansion, and class switching of IgG4$^+$ plasmablasts/plasma cells in a T follicular helper cell (Tfh2)-dependent manner, and the plasmablasts/plasma cells present antigens and activate CD4$^+$ cytotoxic T cells (CTLs), which undergo oligoclonal expansion and drive inflammatory and fibrotic processes that characterize IgG4-RD [3–7]. Therefore, CD4$^+$ T cells and B cells play central roles in the pathogenesis of IgG4-RD.

The vast majority of T cells express αβ T cell receptors (TCRs), which interacts with peptide-MHC complex presented by antigen-presenting cells [8, 9]. During the development of T cells, TCRs are randomly generated through VJ recombination (α chain) or VDJ recombination (β chain), followed by deletion or insertion of non-template nucleotides at junction sites. Then, T cells are subjected to positive and negative selection in the thymus [8, 10]. The diversity of TCRs is predominantly confined to the complementarity-determining regions (CDR). CDR1 and CDR2 domains are encoded by germline V gene segments, while CDR3 domains, the region that directly contacts with peptide antigen, comprise the VJ junction (α chain) or VDJ junction (β chain). Subsequently, CDR3 domains are highly diverse, allowing the recognition of various antigens [11].

The analysis of the TCR repertoire has been challenging due to its enormous diversity. During the past decades, the TCR repertoire was analyzed by CDR3 spectratyping, which involves amplification of CDR3 by RT-PCR using V and J gene-specific primers, and separating the amplicons by polyacrylamide gel electrophoresis [12, 13]. Nowadays, with the advances in high-throughput sequencing technologies, it is possible to sequence millions of TCR clones simultaneously, so that the full TCR repertoire can be analyzed at single-cell resolution, which enables inspection of the adaptive immune system in details [14].

Here, using the approach of high-throughput sequencing, we studied the TCR repertoire of peripheral blood CD4$^+$ T cells from IgG4-RD patients in depth. We compared the characteristics of CD4$^+$ T cell repertoire between healthy controls and IgG4-RD patients, including expansion and coding degeneracy levels of each clonotype, CDR3 length distribution, and usage of TRBV and TRBJ genes. In addition, we analyzed relative similarities of TCR repertoires between individuals, identified IgG4-RD-specific CDR3 amino acid sequences, and clustered the TCR clonotypes based on sequence similarities to reveal disease-specific clusters. Finally, we analyzed the antigen specificities of TCR clonotypes according to the McPAS-TCR database.

Methods

Patients and healthy controls

We included eight newly diagnosed and untreated IgG4-RD patients, and six healthy controls matched for sex, age, and ethnicity. All of the enrolled patients satisfied the 2011 comprehensive diagnostic criteria for definite IgG4-RD [15]. The patients with infectious diseases, malignancies, other rheumatic diseases, or conditions that could mimic IgG4-RD were excluded. All of the enrolled individuals claimed no history of infection or vaccination within 6 months before recruitment. Whole blood samples were collected at Peking Union Medical College Hospital, between May 2017 and Feb 2018. The study protocol was approved by the Ethics Committee of Peking Union Medical College Hospital. All enrolled participants consented to attend this cohort study and signed written informed consent.

CD4$^+$ T cell isolation and RNA extraction

Peripheral blood mononuclear cells (PBMCs) were isolated from fresh blood by standard Ficoll-Hypaque procedures. CD4$^+$ T cells were enriched by positive magnetic-bead selection (Miltenyi, Gladbach, Germany) according to the manufacturer's instructions. RNA was extracted from CD4$^+$ T cells using TRIzol (Invitrogen).

TCR β-chain library preparation and high-throughput sequencing

TCR β-chain sequences were amplified by 5′-rapid amplification of cDNA ends (5′-RACE), using SMARTer® RACE cDNA Amplification Kit (Clontech). The total RNA input was 1 μg. UPM primer was used as 5′ primer, and TRBC-specific primer, with the sequence "TCTGATGGCTCAAACACAGCGACCT," was used as 3′ primer. PCR reaction contained 3 μl cDNA, 20 pmol 5′ and 3′ primers, 4 μl 2.5 mM dNTPs, 2.5 U pfu polymerase, and 10 μl 5 × pfu buffer, with the final volume of 50 μl. PCR conditions were 95 °C for 4 min (min), followed by 25 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 10 s, and a final extension of 72 °C for 5 min. The amplified TCR β-chain products were then cooled to 4 °C. TCR β-chain sequencing libraries were constructed with NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB) and underwent quality control using Bioanalyzer High Sensitivity DNA chip (Agilent). TCR β-chain libraries were sequenced on Illumina Miseq platform (2 × 300 bp).

Bioinformatics analysis

Raw data were processed by Cutadapt software (v1.9.1) [16] to remove adapter sequences and the bases with quality lower than 20. Paired-end reads were merged into one contig sequence by FLASH software (v1.2.11) [17]. Using MiXCR software (v2.0.2) [18], the clean reads were aligned to human TRBV, TRBD, and TRBJ reference sequences, and TCR clonotypes were assembled, with corresponding CDR sequences. TCR repertoire diversity was assessed by the Shannon-Wiener index [9]. The coding degeneracy level was evaluated for each CDR3 amino acid clonotype, calculated as the number of unique nucleotide sequences encoding each CDR3 amino acid sequence [19]. Dimensionality reduction was performed by Barnes-Hut implementation of t-distributed stochastic neighbor embedding (t-SNE) using Rtsne package [20] (1000 iterations, perplexity parameter of 4, trade-off θ of 0.5) and visualized by plotting each event by its t-SNE dimension 1 and dimension 2 in a dot plot.

TCR repertoire similarities between individuals were evaluated by the following metrics using VDJtools [21]: (1) geometric mean of total frequencies of shared clonotypes (metric F), (2) Pearson correlation of frequencies of shared clonotypes (metric R), and (3) normalized number of shared clonotypes (metric D). The repertoire similarities were then visualized by multi-dimensional scaling (MDS) analysis. For TCR network construction, R package "stringdist" [22] was used to calculate Levenshtein distances between each two CDR3 amino acid sequences, and the network figures were made by Cytoscape (http://www.cytoscape.org/) [23]. IgG4-RD-specific clusters were identified in TCR networks. To annotate the TCR clonotypes with known antigen specificity, we referred to the McPAS-TCR database [24]. We annotated all the TCR clonotypes in our dataset of which the Levenshtein distance with the CDR3 sequences in the McPAS-TCR database equals to 0. Detailed information about the calculation of the Shannon-Wiener index, the evaluation of TCR repertoire similarities, the definition of IgG4-RD-specific clusters, and the annotation of TCR clonotypes were provided in Additional file 1. R package "ggplot2" [25], "circlize" [26], and "VennDiagram" [27] were used to plot the figures.

Statistical analysis

Differences between the groups were analyzed by Welch's *t* test for the variables following a normal distribution and the Mann-Whitney *U* test for the variables not following a normal distribution. Correlations between nonnormally distributed variables were analyzed by the Spearman rank correlation test. Outliers identified in this study fulfilled both the Dixon criterion and the Grubbs criterion. Bootstrap resampling was conducted with the following procedures: First, we generated 1000 bootstrap samples by randomly sampling the data from a healthy control or IgG4-RD patients with replacement. Each bootstrap sample had the same sample size as the original dataset. Then, we calculated the mean values in bootstrap samples, analyzed the distribution of bootstrap means according to density plot, and estimated the mean and 95% confidence interval accordingly. The existence of outliers was assessed by bootstrap-based outlier detection plot (Bootlier plot), which is constructed by bootstrapping the statistic "mean-trimmed mean" (MTM) and plotting its distribution [28]. The existence of outliers was indicated by the multimodality of Bootlier plot and tested by the Bootlier test, according to the methodology developed by Candelon and Metiu [29]. We performed a nonparametric bootstrap *t* test with a pooled resampling method for group comparison, according to the recommendation by Dwivedi et al. for small sample size studies [30]. For multiple comparison, false discovery rate (FDR) control was performed by the Benjamini-Hochberg procedure. Data analysis was performed by R

Table 1 Demographic and clinical characteristics of the individuals enrolled in this study

ID	Sex/age	Organ involvement	IgG4-RD RI	Serum IgG4 (mg/L)
HC-1	M/58	–	–	–
HC-2	M/63	–	–	–
HC-3	F/69	–	–	–
HC-4	M/28	–	–	–
HC-5	M/53	–	–	–
HC-6	M/61	–	–	–
PT-1	M/34	Thyroid	4	3450
PT-2	M/61	Salivary glands + retroperitoneum + large artery + lymph nodes + sinus	19	21,000
PT-3	M/56	Lacrimal glands + salivary glands	8	6010
PT-4	M/62	Lacrimal glands + salivary glands + retroperitoneum + lung + large artery + lymph nodes + sinus	25	44,500
PT-5	M/63	Pancreas + prostate + lymph nodes	13	13,600
PT-6	M/48	Pancreas + prostate	12	9300
PT-7	M/61	Lacrimal glands + salivary glands + lung + sinus	16	32,500
PT-8	F/73	Lacrimal glands + ocular adnexa + sinus + retroperitoneum + bladder + lymph nodes	25	23,400

HC healthy control, *PT* IgG4-RD patient, *RI* responder index studio (v3.4.0) and GraphPad Prism 7 software. $p < 0.05$ was considered statistically significant.

Results

The demographic and clinical characteristics of the individuals enrolled in this study were summarized in Table 1. A detailed description of the total number of raw reads, filtered reads, aligned TCR sequences, and unique clonotypes of each sample was displayed in Additional file 2. 244,175 ± 47,618 TCR sequences were obtained from each individual. Based on distinct CDR1, CDR2, and CDR3 nucleotide sequences, 35,099 ± 9319 clonotypes were identified in each sample. The TCR clonotypes identified in each sample were summarized in Additional file 3.

Clonal expansion and coding degeneracy of CD4$^+$ T cells in healthy controls and IgG4-RD patients

To avoid the potential influence of sequencing depth on the analysis of clonotype frequency and diversity,

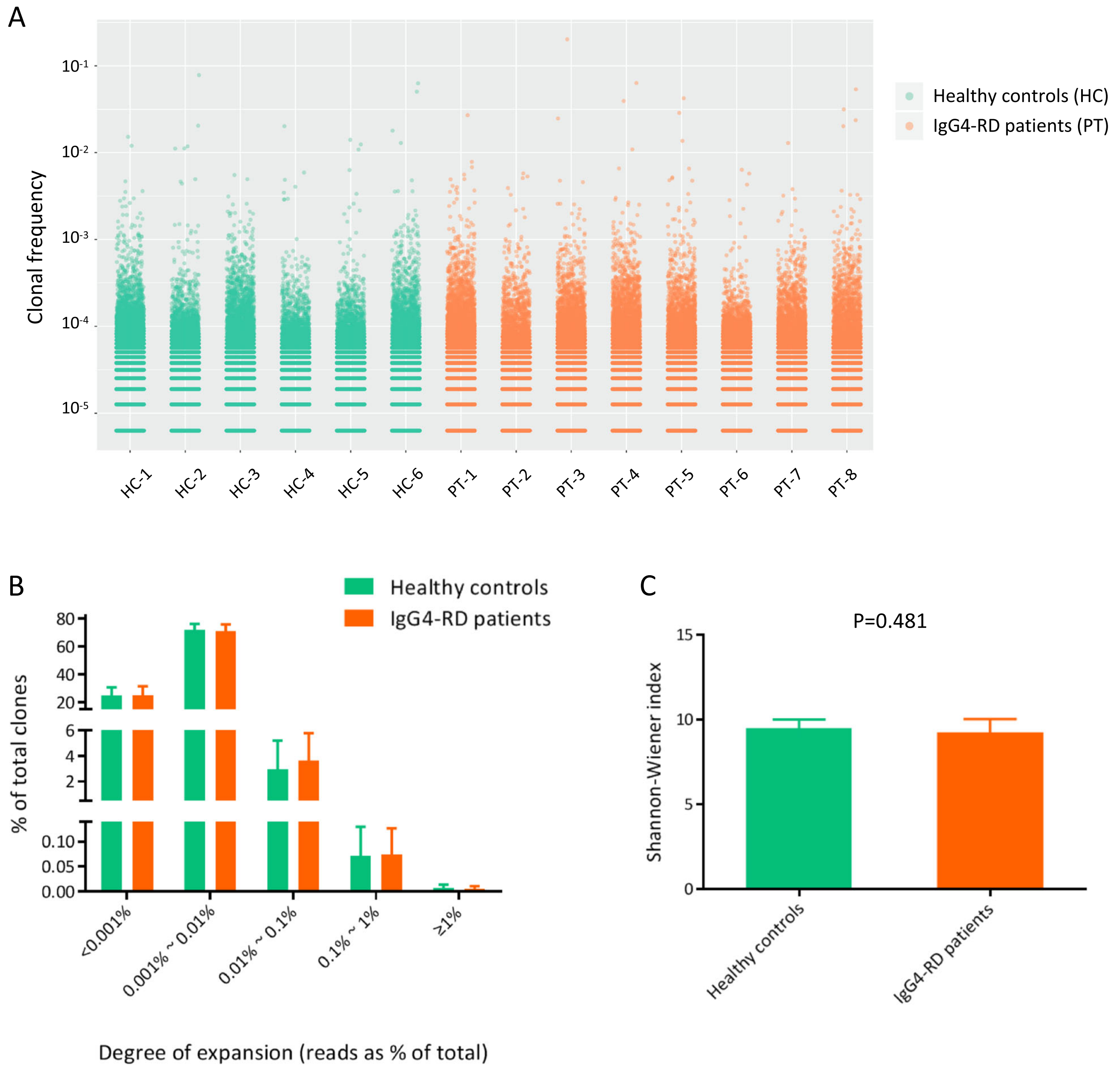

Fig. 1 Clonal expansion and diversity of CD4$^+$ T cell repertoire of healthy controls and IgG4-RD patients. The TCR repertoires of all the individuals were downsampled to the same number of reads (158,416 TCR sequences per sample). **a** Scatterplot showing the frequencies of all the TCR clones in each sample. Each dot represents one clone. **b** Frequency distribution of CD4$^+$ T cell clones in healthy controls and IgG4-RD patients. Data were represented as mean and standard deviation (sd). **c** Shannon-Wiener index as quantification of TCR repertoire diversity of CD4$^+$ T cells. Data were represented as mean and sd. Difference between groups was analyzed by Welch's *t* test

the TCR repertoires of all the individuals involved were downsampled to the same number of reads (158,416 TCR sequences per sample). The degree of expansion of $CD4^+$ T cell clones was assessed by the frequency of clones within a sample. Several clearly expanded $CD4^+$ T cell clones were detected in the background of hundreds of low-frequency clones in each sample (Fig. 1a). In Fig. 1b, the frequencies of $CD4^+$ T cell clones in both healthy controls and IgG4-RD patients showed a right-skewed distribution, in which the majority of clones were of low frequency (< 0.01%). Therefore, we defined the clones with a frequency of ≥ 0.01% to be expanded clones; and the clones with a frequency of ≥ 0.1% to be highly expanded clones. The proportion of both expanded clones and highly expanded clones were comparable between healthy controls and IgG4-RD patients (proportion of expanded clones: healthy controls—3.05% ± 2.28% [outlier included], 2.18% ± 0.92% [HC-1 excluded as outlier]; IgG4-RD patients—3.72% ± 2.19%; Welch's *t* test, $p = 0.589$ [outlier included], $p = 0.108$ [outlier excluded]). The proportion of highly expanded clones: healthy controls, 0.080% ± 0.059%; IgG4-RD patients, 0.081% ± 0.054%; Welch's *t* test, $p = 0.963$). We further evaluated the diversity of $CD4^+$ T cell repertoire of each individual based on the Shannon-Wiener index and found that the TCR repertoire diversity was comparable between healthy controls and IgG4-RD patients (Fig. 1c).

Next, we analyzed the coding degeneracy level of each T cell clonotype, which is a measurement of the number of unique nucleotide sequences encoding a single CDR3 amino acid clonotype, resulting from the degeneracy of genetic code [19]. Here, we also performed analysis on downsampled TCR repertoire data to avoid the bias of sequencing depth. Most of the TCR amino acid clonotypes were encoded by single-nucleotide sequence (Fig. 2a), and the proportion of coding degenerated clones was comparable between healthy controls and IgG4-RD patients (healthy controls, 1.43% ± 0.22%; IgG4-RD patients, 1.5% ± 0.38%; Welch's *t* test, $p = 0.684$).

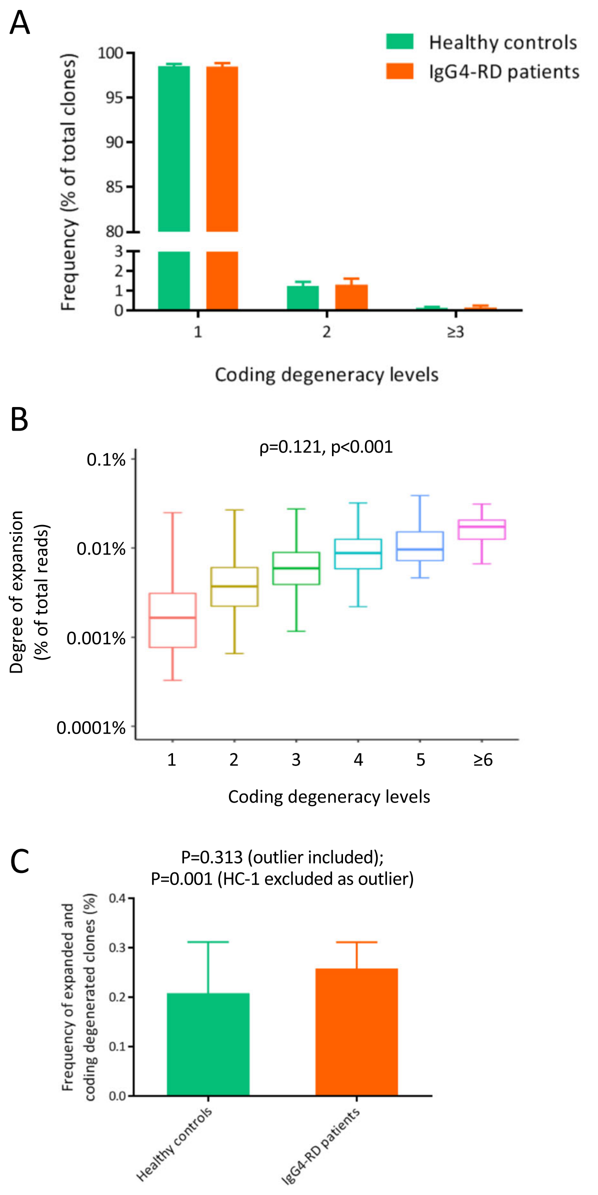

Fig. 2 Coding degeneracy levels of CDR3 amino acid clonotypes. The TCR repertoires of all the individuals were downsampled to the same number of reads (158,416 TCR sequences per sample). **a** Distribution of coding degeneracy levels of CDR3 amino acid clonotypes. Data were represented as mean and sd. **b** Boxplot representing frequencies of CDR3 amino acid clonotypes with each distinct coding degeneracy level. The correlation between expansion levels and coding degeneracy levels of TCR clonotypes were analyzed by Spearman rank correlation test. **c** Proportion of expanded (frequency ≥ 0.01%) and coding degenerated (coding degeneracy level ≥ 2) clones in $CD4^+$ T cell repertoire from healthy controls and IgG4-RD patients. Data were represented as mean and sd. Difference between the groups was analyzed by Welch's *t* test

Based on the fact that the antigen specificity of TCR is determined by the amino acid sequence, the antigen-driven clonal expansion of T cells should increase the frequency of all the TCR nucleotide clonotypes encoding for the same CDR3 amino acid sequence. Therefore, the expanded and coding degenerated clones are likely driven by antigen-specific expansion, while antigen-nonspecific expansion increases only clonal frequencies, not coding degeneracy levels [19]. We observed a positive correlation between coding degeneracy levels and expansion levels of CDR3 amino acid clonotype, according to the pooled data from all 14 individuals (Fig. 2b), suggesting the existence of antigen-driven clonal expansion in $CD4^+$ T cells from peripheral blood. The proportion of expanded (frequency ≥ 0.01%) and coding degenerated

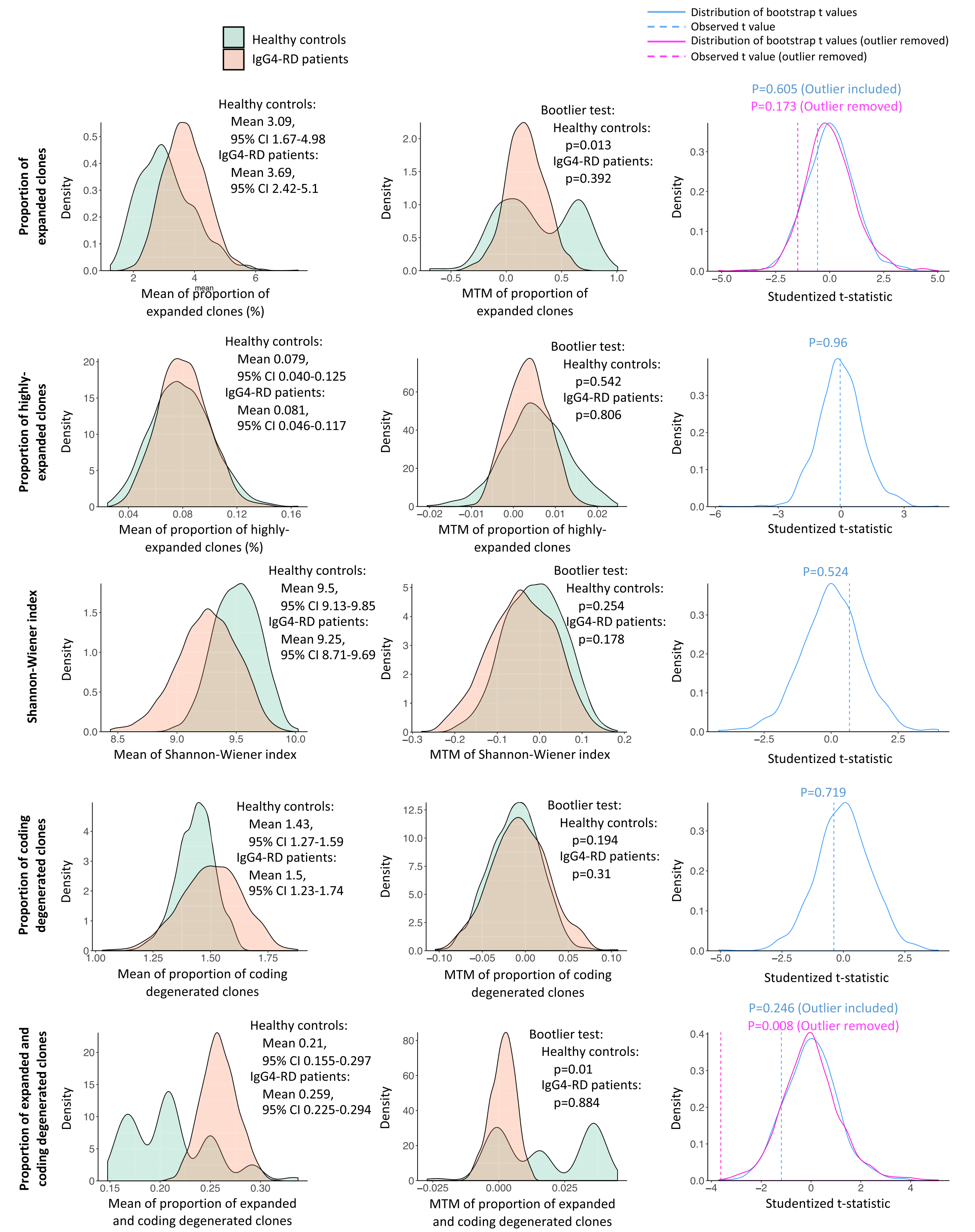

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Bootstrap resampling for the parameters associated with clonal expansion and coding degeneracy levels. Bootstrap resampling was applied to the following parameters: the proportion of expanded clones, the proportion of highly expanded clones, the Shannon-Wiener index, the proportion of coding degenerated clones, and the proportion of expanded and coding degenerated clones. Left panel: The distribution of bootstrap means of each variable in healthy controls and IgG4-RD patients. Estimated means and 95% confidence intervals of each variable were also displayed on the plot. Middle panel: Bootlier plot of each variable in healthy controls and IgG4-RD patients. Results of Bootlier test were also displayed on the plot. Right panel: Group comparison by nonparametric bootstrap *t* test with pooled resampling method. Bootstrap *t* values were calculated according to Dwivedi et al. [30]. The distribution of bootstrap *t* values and observed *t* values was shown

(coding degeneracy level ≥ 2) clones was significantly higher in IgG4-RD patients (Fig. 2c) (healthy controls, 0.209% ± 0.103% [outlier included], 0.167% ± 0.022% [HC-1 excluded as outlier]; IgG4-RD patients, 0.259% ± 0.053%; Welch's *t* test, $p = 0.313$ [outlier included], $p = 0.001$ [outlier excluded]), which suggests the antigen-driven clonal expansion of CD4$^+$ T cells in IgG4-RD.

Bootstrap resampling allowed us to create a large number of simulated datasets from original samples, to make more credible statistical inference without assumptions about the unknown distribution and to model the sampling distribution under a small sample size. We performed bootstrap resampling for the parameters analyzed above, including the proportion of expanded and highly expanded clones, Shannon-Wiener index, the proportion of coding degenerated clones, and the proportion of expanded and coding degenerated clones. The distribution of bootstrap means, as well as estimated means and 95% confidence intervals of each variable, were shown in the left panel of Fig. 3. The outliers identified above (according to the Dixon criterion and the Grubbs criterion) were consistent with the results of the Bootlier plot and Bootlier test (Fig. 3, middle panel). Group comparison was also performed by a nonparametric bootstrap *t* test with a pooled resampling method, which gave similar results as the previous analysis by Welch's *t* test (Fig. 3, right panel).

TRBV and TRBJ gene usage of CD4$^+$ T cells in healthy controls and IgG4-RD patients

The TRBV and TRBJ gene usage of each clonotype was determined using MiXCR. The 100 most frequently used TRBV-TRBJ combinations in healthy controls and IgG4-RD patients were visualized in Fig. 4a, b. We also visualized the patterns of TRBV and TRBJ gene usage of each individual by t-SNE analysis (Fig. 4c–e), which mapped the multi-dimensional data to a two-dimensional space, preserving pairwise similarities of input objects. According to t-SNE maps, most of the healthy controls had similar patterns of TRBV and TRBJ gene expression, while that of IgG4-RD patients was more heterogeneous. To further study the characteristics of TRBV/TRBJ gene usage in IgG4-RD patients, we compared the expression levels of each TRBV and TRBJ gene and the 100 most frequently used TRBV-TRBJ combinations between healthy controls and IgG4-RD patients (Additional files 4, 5, and 6). However, we did not find any significant difference between healthy controls and IgG4-RD patients after false discovery rate (FDR) control.

CDR3 length distribution of CD4$^+$ T cells in IgG4-RD patients was skewed towards longer fragments

We further investigated the CDR3 length distribution of CD4$^+$ T cells from healthy controls and IgG4-RD patients (Fig. 4f). In healthy controls, the length of CDR3 amino acid sequences formed a bell-shaped distribution, which peaks at 14 amino acids. However, the CDR3 length distribution of CD4$^+$ T cells from IgG4-RD patients was skewed towards longer sequences and peaks at 15 amino acids. To further compare the CDR3 length distribution, we downsampled the TCR repertoire of each individual to the same number of reads (158,416 TCR sequences per sample) and combined the sequences from the same group together. Analysis by the Mann-Whitney *U* test revealed a significant difference in CDR3 length distribution between healthy controls and IgG4-RD patients ($p < 0.001$).

Identification of IgG4-RD-specific CDR3 amino acid sequences

Given that it is the amino acid sequence of TCR that determines the structural binding with peptide-MHC complex, the subsequent studies were performed based on the amino acid sequences. A total of 427,682 unique CDR3 amino acid clonotypes of CD4$^+$ T cells were identified from the 14 individuals. Among them, 233,328 clonotypes were found in healthy controls, 176,225 clonotypes were found in IgG4-RD patients, and 18,129 clonotypes were shared between healthy controls and IgG4-RD patients (Fig. 5a). As for the expanded clones (frequency ≥ 0.01%), 7094 clonotypes were found in healthy controls, 4912 clonotypes were found in IgG4-RD patients, and 61 clonotypes were found in both (Fig. 5b).

We evaluated the relative similarity of TCR repertoires between individuals by F, R, and D metrics and built an MDS plot for visualization, as described in the "Methods" section. To avoid the bias of sequencing depth, the analysis was performed on downsampled data.

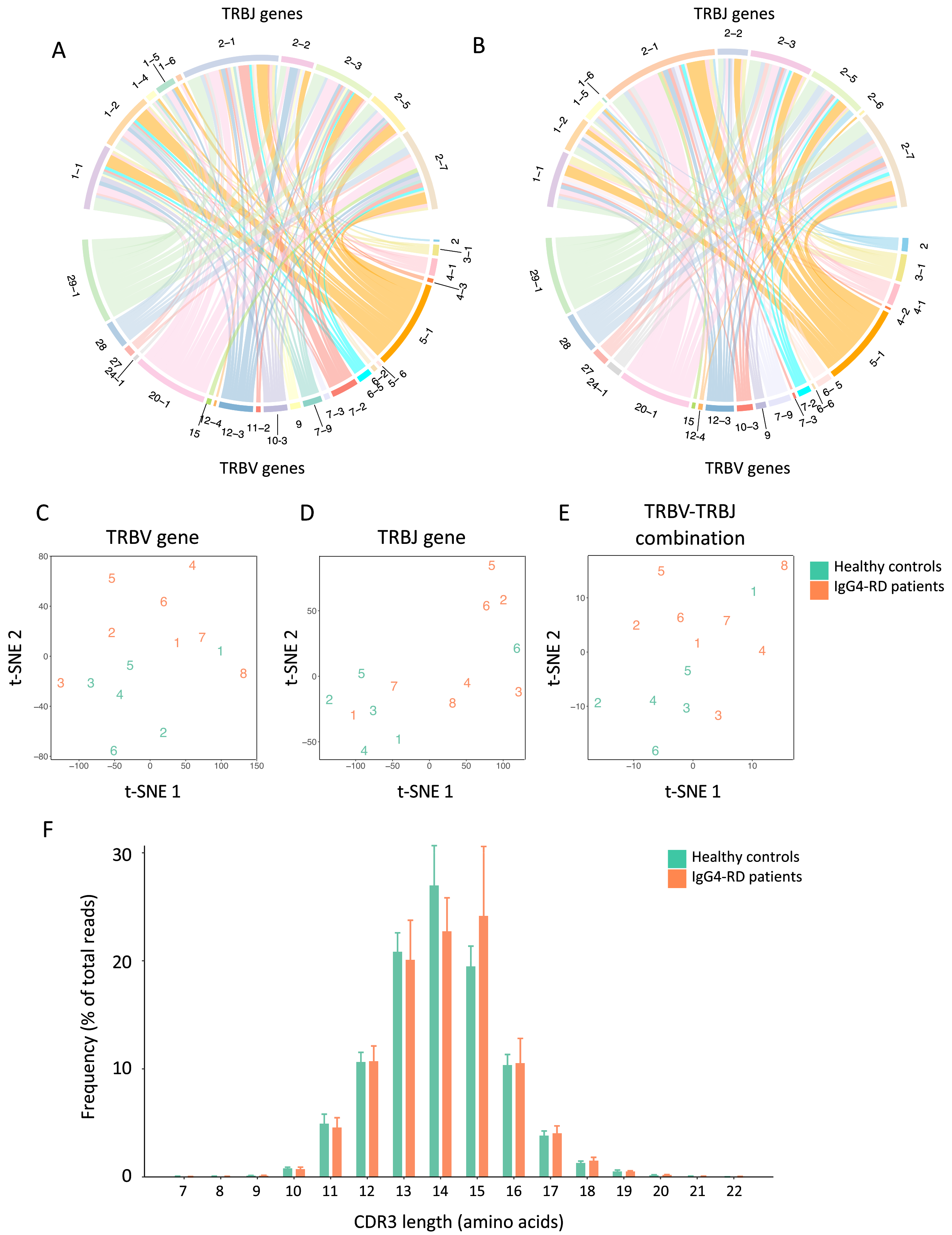

Fig. 4 (See legend on next page.)

(See figure on previous page.)
Fig. 4 TRBV/TRBJ gene usage and CDR3 length distribution. **a** Circos plot representing the 100 most frequently used TRBV-TRBJ gene combinations in CD4$^+$ T cells from healthy controls. **b** Circos plot representing the 100 most frequently used TRBV-TRBJ gene combinations in CD4$^+$ T cells from IgG4-RD patients. **c** Visualization of the patterns of TRBV gene usage of each individual by t-SNE dimensionality reduction analysis. **d** Visualization of the patterns of TRBJ gene usage of each individual by t-SNE dimensionality reduction analysis. **e** Visualization of the patterns of TRBV-TRBJ gene combination of each individual by t-SNE dimensionality reduction analysis. **f** CDR3 length distribution of CD4$^+$ T cells from healthy controls and IgG4-RD patients. Data were represented as mean and sd

According to MDS plots (Fig. 5c–e), TCR repertoires of IgG4-RD patients were separated from that of healthy controls in F and D metrics, while no clear patterns were found in R metrics. Since F and D metrics measure repertoire similarity based on the frequency or number of overlapping sequences, while R metric focuses on the correlation of clonotype frequencies between samples, these data indicated that common clonotypes existed in TCR repertoires among different IgG4-RD patients, but the expansion levels of the shared clonotypes were not correlated in different patients.

To further identify the common TCR clonotypes among IgG4-RD patients, we defined the CDR3 amino acid sequences expanded in at least 2 IgG4-RD patients, while not detected in healthy controls as IgG4-RD-specific sequences, and identified 11 sequences fulfilling these criteria (Table 2).

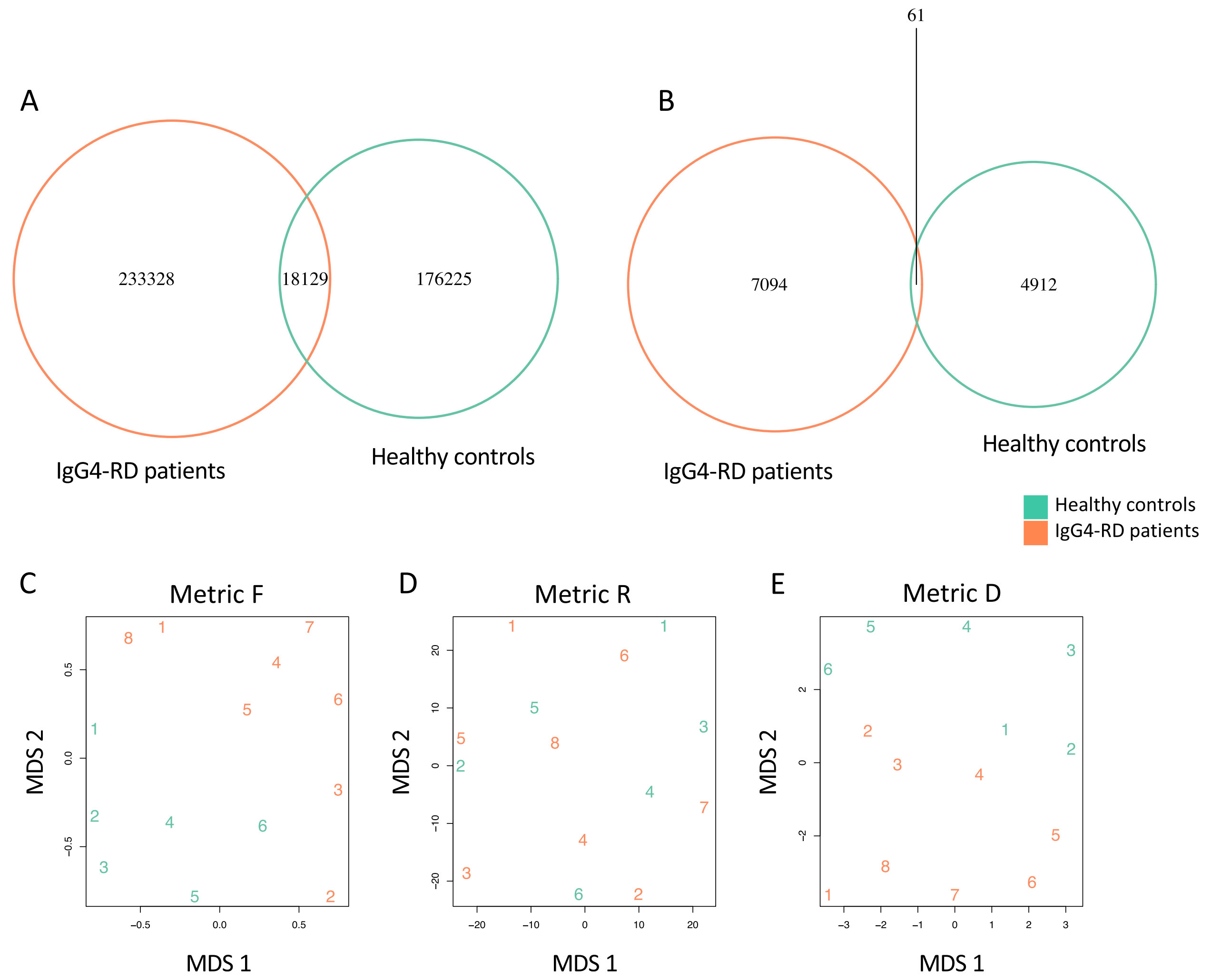

Fig. 5 Relative similarity of CD4$^+$ T cell repertoires between individuals. **a** Venn diagram showing the overlap of CDR3 amino acid clonotypes identified in healthy controls and IgG4-RD patients. **b** Venn diagram showing the overlap of expanded CDR3 amino acid clonotypes identified in healthy controls and IgG4-RD patients. **c**–**e** Multi-dimensional scaling (MDS) plot in which individuals were projected onto two-dimensional space based on pairwise TCR repertoire similarities. Relative similarity of TCR repertoires were evaluated by metric F (**c**), metric R (**d**), and metric D (**e**). Analysis was performed on downsampled data

Of note, the sequence "CASSQGTGVRGTEAFF" was identified in 87.5% of IgG4-RD patients but none of the healthy controls and expanded or highly expanded in 3 patients (frequency 6.33%, 0.03%, and 0.02%, respectively).

Construction of CDR3 similarity networks revealed IgG4-RD-specific clusters

The analysis above did not take into consideration the sequence similarity of different CDR3 clonotypes. Thus, we collected all the CDR3 amino acid clonotypes expanded in at least one individual, evaluated the sequence similarity of each two clonotypes by calculating the Levenshtein distance, and constructed CDR3 similarity networks by connecting each two CDR3 clonotypes (nodes) with a Levenshtein distance of 1 [31]. All the clusters comprised at least four connected nodes as shown in Fig. 6a. There were several large CDR3 clusters shared by healthy controls and IgG4-RD patients, probably reflecting the public sequences in the human TCR repertoire [31]. We also identified several IgG4-RD-specific clusters, and the sequences, source individuals, and expansion levels of each node in IgG4-RD-specific clusters were highlighted in Fig. 6b.

Analysis of TCR sequences with known antigen specificity

In order to reveal the antigen specificity of CD4$^+$ T cell clonotypes, we referred to the McPAS-TCR database, which assembled all the TCR sequences with known antigen specificity based on the published literature and classified the immune functions into 4 categories: pathogens, autoimmune, cancer, and allergy [24]. According to the McPAS-TCR database, we collected 9658 annotated human TCR β-chain CDR3 sequences, of which 1091 annotated sequences were present in the CD4$^+$ T cell repertoire of our dataset. TCR repertoires were downsampled to 158,416 sequences per sample, and the frequencies of sequences associated with each disease were shown in Fig. 7a. Significant expansion of T cell clonotypes associated with influenza or cytomegalovirus (CMV) was found in 3 IgG4-RD patients. However, after false discovery rate control procedures, no significant difference was identified between healthy controls and IgG4-RD. The expression levels of TCR sequences associated with each immune function were also comparable between healthy controls and IgG4-RD patients (Fig. 7b and Additional file 7).

We also looked up the IgG4-RD-specific sequences identified above (Table 2 and Fig. 6b) in the McPAS database. However, none of them was present in the database.

Discussion

The pathogenesis of IgG4-RD has been proposed that Tfh cells induce IgG4 class-switching and differentiation of plasmablasts and plasma cells [6, 32], and B cells and plasmablasts present disease-specific antigens and activate CD4$^+$ CTLs, which drives the inflammatory and fibrotic processes [3–7, 32]. The critical roles of CD4$^+$ T cells in the pathogenesis of IgG4-RD have been demonstrated by this model. Here, using next-generation sequencing, we investigated the TCR repertoire of CD4$^+$ T cells from IgG4-RD patients in-depth, in order to reveal the characteristics of immune repertoire in IgG4-RD.

The clonal expansion of CD4$^+$ T cells was comparable between healthy controls and IgG4-RD patients, revealed by the proportion of TCR clonotypes in each expansion level, as well as the diversity of TCR repertoire calculated by the Shannon-Wiener index (Fig. 1). However, when we took into consideration the coding degeneracy level of each clone, as an indicator of antigen-driven T cell expansion [19], we found a significantly higher proportion of expanded and coding degenerated clones in CD4$^+$ T cells of IgG4-RD patients (Fig. 2), suggesting of antigen-driven clonal expansion in IgG4-RD. Of note, the statistically significant result was based on the removal of the data from HC-1 as an outlier, according to

Table 2 IgG4-RD-specific CDR3 amino acid sequences

CDR3 amino acid sequence	IgG4-RD patients with expanded clones	IgG4-RD patients with clones of lower frequencies
CASSSWTGGRYNSPLHF	PT-1, PT-6	
CATSRRPGLEINNEQFF	PT-1, PT-8	
CASSLSGTNEQFF	PT-1, PT-5	
CASSPPGLDFSGANVLTF	PT-2, PT-7	
CASSQGTGVRGTEAFF	PT-4, PT-5, PT-6	PT-1, PT-3, PT-7, PT-9
CSAPTGGNSGANVLTF	PT-4, PT-7	PT-5
CASSLTGTNTEAFF	PT-4, PT-8	PT-1, PT-3, PT-6
CASSFINEQFF	PT-4, PT-8	
CASSLVSSGSNEQFF	PT-4, PT-5	
CAIGQTYEQYF	PT-4, PT-7	
CATSGFGSEQYF	PT-5, PT-6	PT-1, PT-4, PT-7, PT-8

Listed are the CDR3 amino acid sequences expanded in at least two IgG4-RD patients, while not detected in healthy controls (even at lower frequencies)

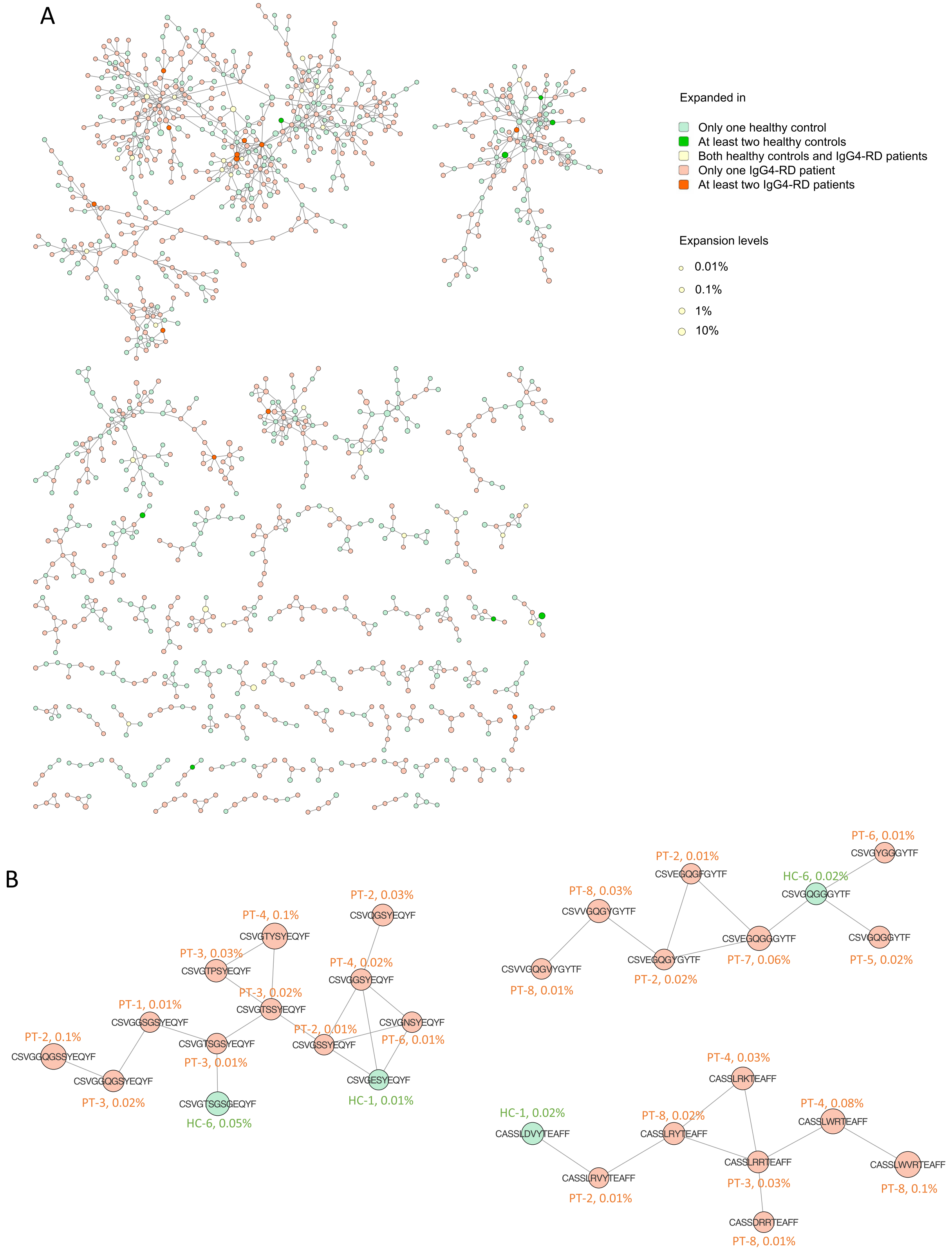

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 IgG4-RD-specific clusters revealed by CDR3 similarity networks. **a** CDR3 similarity networks constructed by all the CDR3 amino acid clonotypes expanded in at least one individual. Each two CDR3 amino acid clonotypes with a Levenshtein distance of 1 were connected with each other. Node sizes correspond to the expansion levels of the clonotypes. If the clonotypes were expressed by more than one individual, the node sizes were determined by the highest expansion level. **b** IgG4-RD-specific clusters and the CDR3 amino acid sequences, source individuals, and expansion levels of each node in the clusters

the Dixon criterion, Grubbs criterion, and Bootlier plot. The TCR repertoire is sensitive to environmental stimulation. Therefore, it could be possible that this healthy donor underwent some subclinical infection or hypersensitivity, resulting in antigen-dependent clonal expansion of the TCR repertoire. However, we are also aware of the potential bias caused by outlier removal.

In contrast to our data, Mattoo et al. reported that the CD4$^+$ CTLs from the peripheral blood of IgG4-RD patients were oligoclonally expanded [7]. The discrepancy of T cell expansion between these two studies could be possibly explained by the fact that the CD4$^+$ CTLs, as a small subpopulation of peripheral CD4$^+$ T cells, did not significantly affect the diversity of CD4$^+$ T cell pool.

It has been revealed that the patterns of TRBV gene usage were determined dominantly by MHC alleles [33]. Given that the human leukocyte antigen (HLA) complex of IgG4-RD patients is biased towards certain susceptible alleles [34, 35], we hypothesized that the TRBV gene usage of TCR repertoire might also be biased in IgG4-RD. However, we did not find any significant difference in TRBV/TRBJ gene usage between healthy controls and IgG4-RD patients (Additional files 4, 5, and 6), and the t-SNE analysis revealed heterogeneity between patients (Fig. 4c–e).

The distribution of CDR3 length was skewed towards longer fragments in CD4$^+$ T cells from IgG4-RD patients (Fig. 4f). A similar pattern of CDR3 length distribution has also been found in peripheral blood T cells from systemic lupus erythematosus patients [36]. It has been reported that antigen exposure resulted in longer CDR3 domains in adults compared to infants in both CD4$^+$ and CD8$^+$ T cell compartments [37], suggesting that longer CDR3 in IgG4-RD patients might be caused by chronic antigen exposure, which is consistent with the previous results showing antigen-driven clonal expansion in IgG4-RD. However, further study on CDR3 length distribution of TCR is required to fully interpret these data.

TCR repertoire dynamically encodes the antigen exposure history of each individual. TCR repertoire signatures have been identified in a variety of disease, including infectious diseases [38–40], autoimmunity [36, 41], and cancer [42, 43], and could serve as a promising biomarker for diagnosis, prognosis, and monitoring of certain diseases [38, 43, 44]. Here, visualization by MDS analysis revealed repertoire similarities among IgG4-RD patients in F and D metrics (Fig. 5c–e), indicating that disease-specific TCR sequence signatures existed in the CD4$^+$ T cell repertoire of IgG4-RD patients, probably driven by common antigens. In addition, we identified the sequence “CASSQGTGVRGTEAFF” (Table 2) that presented in seven of eight IgG4-RD patients but none of the healthy controls. Further study is required to validate the usage of this sequence as a diagnostic biomarker, and to reveal the epitope specificity of this clonotype, thus providing insights for the antigen driving IgG4-RD.

The antigen that triggers IgG4-RD remains unclear. Several antibodies against autoantigens have been identified in IgG4-RD patients, including galectin-3 [45], laminin 511-E8 [46], annexin A11 [47], and prohibitin [48]. However, the positive rate of those antibodies was suboptimal (18~73%), and the antibodies against elements of some organs were undetectable in IgG4-RD patients with other organ involvement [49]. In addition, it is also possible that IgG4-RD could be triggered by the antigens from environmental components, such as pathogens, allergens, and occupational exposure [49–51].

Although of great interest, it is currently impossible to predict the epitopes that T cells recognize based on TCR sequences. However, a recent breakthrough in bioinformatics has allowed us to cluster the TCR sequences with shared epitope specificity [52, 53]. Meysman et al. compared the available approaches of unsupervised TCR clustering based on several different algorithms of CDR3 similarity assessment and reported that compared to more complicated methods, clustering TCR sequences by simply connecting those with a Levenshtein distance of 1 was already of high performance [54]. Here, we constructed TCR networks based on Levenshtein distance and identified several IgG4-RD-specific clusters (Fig. 6), which provides clues for disease-specific antigens.

Another approach of analyzing the antigen specificity of T cell clonotypes is referring to TCR sequence databases [55]. Here, we referred to the McPAS-TCR database, which not only assembled all the TCR sequences with known antigen specificities, but also annotated each sequence according to immune functions. However, no significant difference was found between healthy controls and IgG4-RD patients in the frequency of sequences associated with each disease or immune function (Fig. 7), which could possibly result from the fact that the McPAS database collected only a small fraction of sequences in human TCR repertoire, without the inclusion of IgG4-RD-specific sequences. Of note, although not reaching

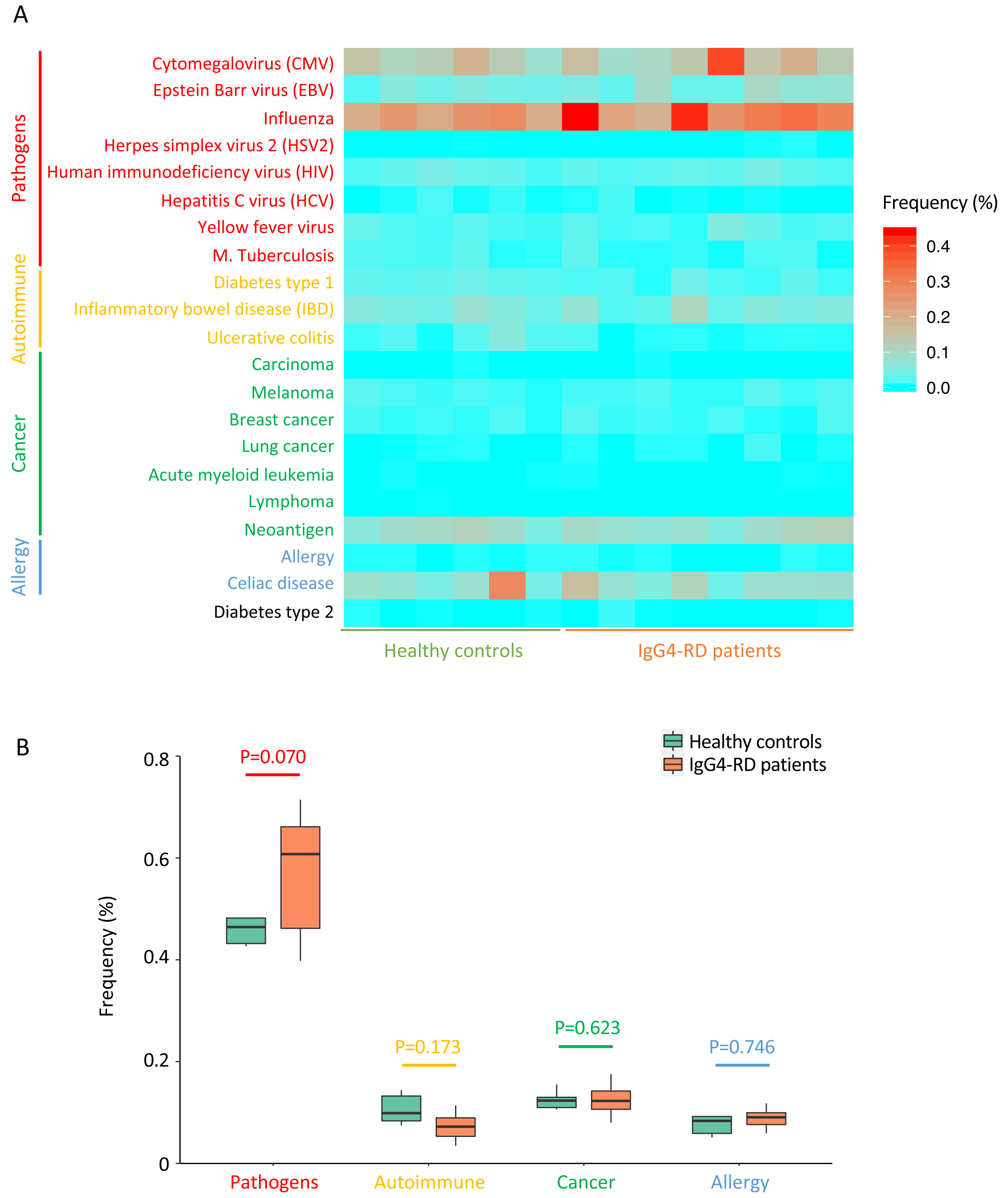

Fig. 7 Expression levels of CDR3 sequences associated with known antigen specificity according to McPAS-TCR database. **a** Heatmap presenting the frequencies of sequences associated with each disease in CD4$^+$ T cell repertoire of each individual. **b** Boxplot summarizing the frequencies of sequences associated with each immune function in CD4$^+$ T cell repertoire of healthy controls and IgG4-RD patients. Difference between the groups was analyzed by Welch's *t* test (for the frequencies of sequences associated with pathogens, autoimmune, and cancer) or Mann-Whitney *U* test (for the frequencies of sequences associated with allergy). Analysis was performed on downsampled data

statistical significance, $CD4^+$ T cells from IgG4-RD patients expressed a relatively higher frequency of TCR sequences associated with pathogens (Fig. 7b), suggesting the underlying association between pathogen exposure and IgG4-RD. However, further study is required to confirm this hypothesis.

Conclusions

In conclusion, using high-throughput sequencing, we analyzed the TCR repertoire of peripheral blood $CD4^+$ T cells in IgG4-RD patients in-depth, and found that the TCR repertoire diversity was comparable between healthy controls and IgG4-RD patients, while there was significantly more expanded and coding degenerated clones in $CD4^+$ T cell repertoire of IgG4-RD patients, suggesting antigen-driven clonal expansion. The CDR3 length distribution of IgG4-RD patients was skewed towards longer fragments. In addition, IgG4-RD-specific CDR3 sequences and clusters were identified in our study, which provides clues for the disease-specific antigen.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13075-019-2069-6.

Additional file 1. : Supplementary methods.

Additional file 2. : TCRβ sequence statistics.

Additional file 3. : Summary of TCR clonotypes in each sample. (CSV 65491 kb)

Additional file 4. : Comparison of TRBV gene usage between healthy controls and IgG4-RD patients.

Additional file 5. : Comparison of TRBJ gene usage between healthy controls and IgG4-RD patients.

Additional file 6. : Comparison of the 100 most frequently-used TRBV-TRBJ combinations between healthy controls and IgG4-RD patients.

Additional file 7. : Bootstrap resampling for the frequency of sequences associated with each immune function. Bootstrap resampling was applied to the following parameters: the frequency of TCR sequences associated with pathogens, autoimmune, cancer, and allergy. Left panel: The distribution of bootstrap means of each variable in healthy controls and IgG4-RD patients. Estimated means and 95% confidence intervals of each variable were also displayed on the plot. Middle panel: Bootlier plot of each variable in healthy controls and IgG4-RD patients. Results of Bootlier test were also displayed on the plot. Right panel: Group comparison by nonparametric bootstrap t-test with pooled resampling method. Bootstrap t values were calculated according to Dwivedi et al. [30]. The distribution of bootstrap t values and observed t values were shown.

Abbreviations

IgG4-RD: IgG4-related disease; 5′-RACE: 5′-Rapid amplification of cDNA ends; CDR3: Complementarity determining region 3; Tfh: T follicular helper cell; CTL: Cytotoxic T lymphocytes; TCR: T cell receptor; PBMC: Peripheral blood mononuclear cell; t-SNE: t-distributed stochastic neighbor embedding; MDS: Multi-dimensional scaling; MTM: Mean-trimmed mean; FDR: False discovery rate; CMV: Cytomegalovirus; HLA: Human leukocyte antigen

Acknowledgements

Not applicable.

Authors' contribution

LW designed the research, analyzed the data, and wrote the manuscript. PZ performed the experiments, collected the data, and revised the manuscript. JLi, HL, LP, and JLing participated in the sample and data collection. XZhang and XZeng helped optimize the research and proofread the paper. YZ and WZ designed the study and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Chinese Academy of Medical Sciences Initiative for Innovative Medicine (2017-12M-3-001), the National Natural Science Foundation of China (No. 81571587, 81771757), and the National Key Research and Development Program of China (No. 2016YFC0901500).

Availability of data and materials

The sequencing data from this study will be made freely available from the NCBI Short Read Archive (SRA).

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of Peking Union Medical College Hospital. All enrolled patients consented to attend this cohort study and signed a written informed consent.

Consent for publication

Yes, we obtained consent for publication from all the individuals of whom detailed information was included in the manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Rheumatology, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Key Laboratory of Rheumatology and Clinical Immunology, Ministry of Education, No.41 Da Mu Cang, Western District, Beijing 100032, People's Republic of China. [2]Department of General Surgery, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. [3]Tsinghua University School of Medicine, Beijing, China.

Received: 20 May 2019 Accepted: 25 November 2019
Published online: 19 December 2019

References

1. Kamisawa T, Zen Y, Pillai S, Stone JH. IgG4-related disease. Lancet (London). 2015;385(9976):1460–71.
2. Miyabe K, Zen Y, Cornell LD, Rajagopalan G, Chowdhary VR, Roberts LR, Chari ST. Gastrointestinal and extra-intestinal manifestations of IgG4-related disease. Gastroenterology. 2018;155(4):990–1003.e1001.
3. Mattoo H, Stone JH, Pillai S. Clonally expanded cytotoxic CD4(+) T cells and the pathogenesis of IgG4-related disease. Autoimmunity. 2017;50(1):19–24.
4. Baptista B, Casian A, Gunawardena H, D'Cruz D, Rice CM. Neurological manifestations of IgG4-related disease. Curr Treat Options Neurol. 2017;19(4):14.
5. Mattoo H, Mahajan VS, Della-Torre E, Sekigami Y, Carruthers M, Wallace ZS, Deshpande V, Stone JH, Pillai S. De novo oligoclonal expansions of circulating plasmablasts in active and relapsing IgG4-related disease. J Allergy Clin Immunol. 2014;134(3):679–87.
6. Chen Y, Lin W, Yang H, Wang M, Zhang P, Feng R, Chen H, Peng L, Zhang X, Zhao Y, et al. Aberrant expansion and function of follicular helper T cell subsets in IgG4-related disease. Arthritis Rheumatol. 2018;70(11):1853–65.
7. Mattoo H, Mahajan VS, Maehara T, Deshpande V, Della-Torre E, Wallace ZS, Kulikova M, Drijvers JM, Daccache J, Carruthers MN, et al. Clonal expansion of CD4(+) cytotoxic T lymphocytes in patients with IgG4-related disease. J Allergy Clin Immunol. 2016;138(3):825–38.
8. Woodsworth DJ, Castellarin M, Holt RA. Sequence analysis of T-cell repertoires in health and disease. Genome Med. 2013;5(10):98.
9. Hou XL, Wang L, Ding YL, Xie Q, Diao HY. Current status and recent advances of next generation sequencing techniques in immunological repertoire. Genes Immun. 2016;17(3):153–64.
10. Richards DM, Kyewski B, Feuerer M. Re-examining the nature and function of self-reactive T cells. Trends Immunol. 2016;37(2):114–25.

11. Attaf M, Huseby E, Sewell AK. αβ T cell receptors as predictors of health and disease. Cell Mol Immunol. 2015;12(4):391–9.
12. Pannetier C, Cochet M, Darche S, Casrouge A, Zoller M, Kourilsky P. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. Proc Natl Acad Sci U S A. 1993;90(9):4319–23.
13. Gorski J, Yassai M, Zhu X, Kissela B, Kissella B, Keever C, Flomenberg N. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. J Immunol. 1994;152(10):5109–19.
14. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. Genome Res. 2009;19(10):1817–24.
15. Umehara H, Okazaki K, Masaki Y, Kawano M, Yamamoto M, Saeki T, Matsui S, Yoshino T, Nakamura S, Kawa S, et al. Comprehensive diagnostic criteria for IgG4-related disease (IgG4-RD), 2011. Mod Rheumatol. 2012;22(1):21–30.
16. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1):10–2.
17. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957–63.
18. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12(5):380–1.
19. Jia Q, Zhou J, Chen G, Shi Y, Yu H, Guan P, Lin R, Jiang N, Yu P, Li QJ, et al. Diversity index of mucosal resident T lymphocyte repertoire predicts clinical prognosis in gastric cancer. Oncoimmunology. 2015;4(4):e1001230.
20. Krijthe J. Rtsne: t-distributed stochastic neighbor embedding using Barnes-Hut implementation; 2015.
21. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI, et al. VDJtools: unifying post-analysis of T cell receptor repertoires. PLoS Comput Biol. 2015;11(11):e1004503.
22. van der Loo M. The stringdist package for approximate string matching. R J. 2014;6(1):111–22.
23. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, et al. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007;2(10):2366–82.
24. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics (Oxford). 2017;33(18):2924–9.
25. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
26. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. Bioinformatics (Oxford). 2014;30(19):2811–2.
27. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
28. Singh K, Xie M. Bootlier-plot: bootstrap based outlier detection plot. Sankhyā. 2003;65(3):532–59.
29. Candelon B, Metiu N: A distribution-free test for outliers. Bundesbank Discussion Paper No 02/2013.
30. Dwivedi AK, Mallawaarachchi I, Alvarado LA. Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. Stat Med. 2017;36(14):2187–205.
31. Madi A, Poran A, Shifrut E, Reich-Zeliger S, Greenstein E, Zaretsky I, Arnon T, Laethem FV, Singer A, Lu J, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. eLife. 2017;6:e22057.
32. Akiyama M, Suzuki K, Yasuoka H, Kaneko Y, Yamaoka K, Takeuchi T. Follicular helper T cells in the pathogenesis of IgG4-related disease. Rheumatology (Oxford). 2018;57(2):236–45.
33. Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, Chain B, Cohen IR, Friedman N. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. Genome Res. 2014;24(10):1603–12.
34. Ota M, Umemura T, Kawa S. Immunogenetics of IgG4-related AIP. Curr Top Microbiol Immunol. 2017;401:35–44.
35. Stone JH, Zen Y, Deshpande V. IgG4-related disease. N Engl J Med. 2012;366(6):539–51.
36. Sui W, Hou X, Zou G, Che W, Yang M, Zheng C, Liu F, Chen P, Wei X, Lai L, et al. Composition and variation analysis of the TCR beta-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. Mol Immunol. 2015;67(2 Pt B):455–64.
37. Hall MA, Reid JL, Lanchbury JS. The distribution of human TCR junctional region lengths shifts with age in both CD4 and CD8 T cells. Int Immunol. 1998;10(10):1407–19.
38. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, Desmarais C, Klinger M, Carlson CS, Hansen JA, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat Genet. 2017;49(5):659–65.
39. De Neuter N, Bartholomeus E, Elias G, Keersmaekers N, Suls A, Jansens H, Smits E, Hens N, Beutels P, Van Damme P, et al. Memory CD4(+) T cell receptor repertoire data mining as a tool for identifying cytomegalovirus serostatus. Genes Immun. 2019;20(3):255–60.
40. Jiang Q, Liu Y, Xu B, Zheng W, Xiang X, Tang X, Dong H, Chen Y, Wang C, Deng G, et al. Analysis of T cell receptor repertoire in monozygotic twins concordant and discordant for chronic hepatitis B infection. Biochem Biophys Res Commun. 2018;497(1):153–9.
41. Doorenspleet ME, Westera L, Peters CP, Hakvoort TBM, Esveldt RE, Vogels E, van Kampen AHC, Baas F, Buskens C, Bemelman WA, et al. Profoundly expanded T-cell clones in the inflamed and uninflamed intestine of patients with Crohn's disease. J Crohn's Colitis. 2017;11(7):831–9.
42. Ostmeyer J, Christley S, Toby IT, Cowell LG. Biophysicochemical motifs in T-cell receptor sequences distinguish repertoires from tumor-infiltrating lymphocyte and adjacent healthy tissue. Cancer Res. 2019;79(7):1671–80.
43. Cui JH, Lin KR, Yuan SH, Jin YB, Chen XP, Su XK, Jiang J, Pan YM, Mao SL, Mao XF, et al. TCR repertoire as a novel indicator for immune monitoring and prognosis assessment of patients with cervical cancer. Front Immunol. 2018;9:2729.
44. Wu D, Emerson RO, Sherwood A, Loh ML, Angiolillo A, Howie B, Vogt J, Rieder M, Kirsch I, Carlson C, et al. Detection of minimal residual disease in B lymphoblastic leukemia by high-throughput sequencing of IGH. Clin Cancer Res. 2014;20(17):4540–8.
45. Perugino CA, AlSalem SB, Mattoo H, Della-Torre E, Mahajan V, Ganesh G, Allard-Chamard H, Wallace Z, Montesi SB, Kreuzer J, et al. Identification of galectin-3 as an autoantigen in patients with IgG4-related disease. J Allergy Clin Immunol. 2019;143(2):736-45.e736.
46. Shiokawa M, Kodama Y, Sekiguchi K, Kuwada T, Tomono T, Kuriyama K, Yamazaki H, Morita T, Marui S, Sogabe Y, et al. Laminin 511 is a target antigen in autoimmune pancreatitis. Sci Transl Med. 2018;10(453):eaaq0997.
47. Hubers LM, Vos H, Schuurman AR, Erken R, Oude Elferink RP, Burgering B, van de Graaf SFJ, Beuers U. Annexin A11 is targeted by IgG4 and IgG1 autoantibodies in IgG4-related disease. Gut. 2018;67(4):728–35.
48. Du H, Shi L, Chen P, Yang W, Xun Y, Yang C, Zhao L, Zhou Y, Chen G. Prohibitin is involved in patients with IgG4 related disease. PLoS One. 2015;10(5):e0125331.
49. Umehara H, Okazaki K, Kawano M, Tanaka Y. The front line of research into immunoglobin (Ig) G4-related disease - do autoantibodies cause IgG4-RD? Mod Rheumatol. 2019;29(2):214-8.
50. de Buy Wenniger LJ, Culver EL, Beuers U. Exposure to occupational antigens might predispose to IgG4-related disease. Hepatology (Baltimore). 2014;60(4):1453–4.
51. Culver EL, Sadler R, Bateman AC, Makuch M, Cargill T, Ferry B, Aalberse R, Barnes E, Rispens T. Increases in IgE, eosinophils, and mast cells can be used in diagnosis and to predict relapse of IgG4-related disease. Clin Gastroenterol Hepatol. 2017;15(9):1444–1452.e1446.
52. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, Crawford JC, Clemens EB, Nguyen THO, Kedzierska K, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature. 2017;547(7661):89–93.
53. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, Ji X, Han A, Krams SM, Pettus C, et al. Identifying specificity groups in the T cell receptor repertoire. Nature. 2017;547(7661):94–8.
54. Meysman P, De Neuter N, Gielis S, Bui Thi D, Ogunjimi B, Laukens K. On the viability of unsupervised T-cell receptor sequence clustering for epitope preference. Bioinformatics (Oxford). 2019;35(9):1461-8.
55. Heather JM, Ismail M, Oakes T, Chain B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform. 2018;19(4):554–65.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.